

106

**PCT**

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



( INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/00, 15/09, 15/29, 15/82, A01H</b> <b>1/00, 3/00, 5/00</b>		<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/37184</b>  <b>(43) International Publication Date:</b> 27 August 1998 (27.08.98)
<b>(21) International Application Number:</b> PCT/US98/02998 <b>(22) International Filing Date:</b> 20 February 1998 (20.02.98)  <b>(30) Priority Data:</b> 08/804,534 21 February 1997 (21.02.97) US Not furnished 19 February 1998 (19.02.98) US  <b>(71) Applicant:</b> THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [-/US]; 22nd floor, 300 Lakeside Drive, Oakland, CA 94612-3550 (US).  <b>(72) Inventors:</b> HARADA, John, J.; 115 Grande Avenue, Davis, CA 95616 (US). LOTAN, Tamar; 408 Russell Park #4, Davis, CA 95616 (US). OHTO, Masa-aki; 2-5-1, Tatsumi-minami #1-23, Okazaki 444 (JP). GOLDBERG, Robert, B.; 21907 Canon Drive, Topanga, CA 90290 (US). FISCHER, Robert, L.; 1423 Scott Street, El Cerrito, CA 94530 (US).  <b>(74) Agent:</b> BERLINER, Robert; Fulbright & Jaworski, L.L.P., 29th floor, 865 S. Figueroa Street, Los Angeles, CA 90017-2571 (US).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i>	
<b>(54) Title:</b> LEAFY COTYLEDON1 GENES AND THEIR USES			
<b>(57) Abstract</b>  The present invention provides nucleic acid sequences from embryo-specific genes. The nucleic acids are useful in targeting gene expression to embryos or in modulating embryo development.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## LEAFY COTYLEDON1 GENES AND THEIR USES

## FIELD OF THE INVENTION

5           The present invention is directed to plant genetic engineering. In particular, it relates to new embryo-specific genes useful in improving agronomically important plants.

## BACKGROUND OF THE INVENTION

10           Embryogenesis in higher plants is a critical stage of the plant life cycle in which the primary organs are established. Embryo development can be separated into two main phases: the early phase in which the primary body organization of the embryo is laid down and the late phase which involves maturation, desiccation and dormancy. In the early phase, the symmetry of the embryo changes from radial to bilateral, giving rise to a hypocotyl with a  
15           shoot meristem surrounded by the two cotyledonary primordia at the apical pole and a root meristem at the basal pole. In the late phase, during maturation the embryo achieves its maximum size and the seed accumulates storage proteins and lipids. Maturation is ended by the desiccation stage in which the seed water content decreases rapidly and the embryo passes into metabolic quiescent state. Dormancy ends with seed germination, and development  
20           continues from the shoot and the root meristem regions.

          The precise regulatory mechanisms which control cell and organ differentiation during the initial phase of embryogenesis are largely unknown. The plant hormone abscisic acid (ABA) is thought to play a role during late embryogenesis, mainly in the maturation stage by inhibiting germination during embryogenesis (Black, M. (1991). In *Absciscic Acid: Physiology and Biochemistry*, W. J. Davies and H. G. Jones, eds. (Oxford: Bios Scientific Publishers Ltd.), pp. 99-124) Koornneef, M., and Karssen, C. M. (1994). In *Arabidopsis*, E. M. Meyerowitz and C. R. Somerville, eds. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 313-334). Mutations which effect seed development and are ABA insensitive have been identified in *Arabidopsis* and maize. The ABA insensitive (*abi3*) mutant  
25           of *Arabidopsis* and the viviparous1 (*vp1*) mutant of maize are detected mainly during late embryogenesis (McCarty, *et al.*, (1989) *Plant Cell* 1, 523-532 and Parcy *et al.*, (1994) *Plant Cell* 6, 1567-1582). Both the *VP1* gene and the *ABI3* genes have been isolated and were  
30

found to share conserved regions (Giraudat, J. (1995) *Current Opinion in Cell Biology* 7:232-238 and McCarty, D. R. (1995). *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 46:71-93). The *VPI* gene has been shown to function as a transcription activator (McCarty, *et al.*, (1991) *Cell* 66:895-906). It has been suggested that *ABI3* has a similar function.

Another class of embryo defective mutants involves three genes: *LEAFY*, *COTYLEDON1* and 2 (*LEC1*, *LEC2*) and *FUSCA3* (*FUS3*). These genes are thought to play a central role in late embryogenesis (Baumlein, *et al.* (1994) *Plant J.* 6:379-387; Meinke, D. W. (1992) *Science* 258:1647-1650; Meinke *et al.*, *Plant Cell* 6:1049-1064; West *et al.*, (1994) *Plant Cell* 6:1731-1745). Like the *abi3* mutant, leafy cotyledon-type mutants are defective in late embryogenesis. In these mutants, seed morphology is altered, the shoot meristem is activated early, storage proteins are lacking and developing cotyledons accumulate anthocyanin. As with *abi3* mutants, they are desiccation intolerant and therefore die during late embryogenesis. Nevertheless, the immature mutants embryos can be rescued to give rise to mature and fertile plants. However, unlike *abi3* when the immature mutants germinate they exhibit trichomes on the adaxial surface of the cotyledon. Trichomes are normally present only on leaves, stems and sepals, not cotyledons. Therefore, it is thought that the leafy cotyledon type genes have a role in specifying cotyledon identity during embryo development.

Among the above mutants, the *lec1* mutant exhibits the most extreme phenotype during embryogenesis. For example, the maturation and postgermination programs are active simultaneously in the *lec1* mutant (West *et al.*, 1994), suggesting a critical role for *LEC1* in gene regulation during late embryogenesis.

In spite of the recent progress in defining the genetic control of embryo development, further progress is required in the identification and analysis of genes expressed specifically in the embryo and seed. Characterization of such genes would allow for the genetic engineering plants with a variety of desirable traits. For instance, modulation of the expression of genes which control embryo development may be used to alter traits such as accumulation of storage proteins in leaves and cotyledons. Alternatively, promoters from embryo or seed-specific genes can be used to direct expression of desirable heterologous genes to the embryo or seed. The present invention addresses these and other needs.

## SUMMARY OF THE INVENTION

The present invention is based, in part, on the isolation and characterization of *LEC1* genes. The invention provides isolated nucleic acid molecules comprising a *LEC1* polynucleotide sequence, typically about 630 nucleotides in length, which specifically  
5 hybridizes to SEQ. ID. No. 1 under stringent conditions. The *LEC1* polynucleotides of the invention can encode a LEC1 polypeptide of about 210 amino acids, typically as shown in SEQ. ID. No. 2.

The nucleic acids of the invention may also comprise expression cassettes containing a plant promoter operably linked to the *LEC1* polynucleotide. In some  
10 embodiments, the promoter is from a *LEC1* gene, for instance, as shown in SEQ. ID. No. 3. The *LEC1* polynucleotide may be linked to the promoter in a sense or antisense orientation.

The invention also provides transgenic plants comprising an expression cassette containing a plant promoter operably linked to a heterologous *LEC1* polynucleotide. The *LEC1* may encode a LEC1 polypeptide or may be linked to the  
15 promoter in an antisense orientation. The plant promoter may be from any number of sources, including a *LEC1* gene, such as that shown in SEQ. ID. No. 3 or SEQ. ID. No. 4. The transgenic plant can be any desired plant but is often a member of the genus *Brassica*.

Methods of modulating seed development in a plants are also provided. The methods comprise introducing into a plant an expression cassette containing a plant promoter  
20 operably linked to a heterologous *LEC1* polynucleotide. The *LEC1* may encode a LEC1 polypeptide or may be linked to the promoter in an antisense orientation. The expression cassette can be introduced into the plant by any number of means known in the art, including through a sexual cross.

The invention further provides expression cassettes containing promoter  
25 sequences from *LEC1* genes. The promoters of the invention can be characterized by their ability to specifically hybridizes to a polynucleotide sequence consisting of nucleotides 1 to - 1998 of SEQ. ID. No. 3. The promoters of the invention can be operably linked to a variety of nucleic acids, whose expression is to be targeted to embryos or seeds. Transgenic plants comprising the expression cassettes are also provided.

30 The promoters of the invention can be used in methods of targeting expression of a desired polynucleotide to seeds. The methods comprise introducing into a plant an

expression cassette containing a *LEC1* promoter operably linked to a heterologous polynucleotide sequence.

#### Definitions

5           The phrase "nucleic acid" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Nucleic acids may also include modified nucleotides that permit correct read through by a polymerase and do not alter expression of a polypeptide encoded by that nucleic acid.

10           The phrase "polynucleotide sequence" or "nucleic acid sequence" includes both the sense and antisense strands as either individual single strands or in the duplex. It includes, but is not limited to, self-replicating plasmids, chromosomal sequences, and infectious polymers of DNA or RNA.

15           The phrase "nucleic acid sequence encoding" refers to a nucleic acid which directs the expression of a specific protein or peptide. The nucleic acid sequences include both the DNA strand sequence that is transcribed into RNA and the RNA sequence that is translated into protein. The nucleic acid sequences include both the full length nucleic acid sequences as well as non-full length sequences derived from the full length sequences. It should be further understood that the sequence includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in a  
20           specific host cell.

          The term "promoter" refers to a region or sequence determinants located upstream or downstream from the start of transcription and which are involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "plant promoter" is a promoter capable of initiating transcription in plant cells. Such  
25           promoters need not be of plant origin, for example, promoters derived from plant viruses, such as the CaMV35S promoter, can be used in the present invention.

          The term "plant" includes whole plants, plant organs (e.g., leaves, stems, flowers, roots, etc.), seeds and plant cells and progeny of same. The class of plants which can be used in the method of the invention is generally as broad as the class of higher  
30           plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants, as well as certain lower plants such as algae. It includes plants of a variety of ploidy levels, including polyploid, diploid and haploid.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants. As defined here, a modified *LEC1* coding sequence which is heterologous to an operably linked *LEC1* promoter does not include the T-DNA insertional mutants as described in West *et al.*, *The Plant Cell* 6:1731-1745 (1994).

A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, *in planta* techniques, and the like. Such a plant containing the exogenous nucleic acid is referred to here as an R<sub>1</sub> generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are descendants of such a plant.

As used herein an "embryo-specific gene" or "seed specific gene" is a gene that is preferentially expressed during embryo development in a plant. For purposes of this disclosure, embryo development begins with the first cell divisions in the zygote and continues through the late phase of embryo development (characterized by maturation, desiccation, dormancy), and ends with the production of a mature and desiccated seed. Embryo-specific genes can be further classified as "early phase-specific" and "late phase-specific". Early phase-specific genes are those expressed in embryos up to the end of embryo morphogenesis. Late phase-specific genes are those expressed from maturation through to production of a mature and desiccated seed.

A "*LEC1* polynucleotide" is a nucleic acid sequence comprising (or consisting of) a coding region of about 100 to about 900 nucleotides, sometimes from about 300 to about 630 nucleotides, which hybridizes to SEQ. ID. No. 1 under stringent conditions (as defined below), or which encodes a LEC1 polypeptide. *LEC1* polynucleotides can also be identified by their ability to hybridize under low stringency conditions (*e.g.*, T<sub>m</sub> -40°C) to nucleic acid probes having a sequence from position 1 to 81 in SEQ. ID. NO. 1 or from position 355 to 627 in SEQ. ID. NO. 1.

A "promoter from a *LEC1* gene" or "*LEC1* promoter" will typically be about 500 to about 2000 nucleotides in length, usually from about 750 to 1500. An exemplary promoter sequence is shown as nucleotides 1-1998 of SEQ. ID. No. 3. A *LEC1* promoter can also be identified by its ability to direct expression in all, or essentially all, proglobular embryonic cells, as well as cotyledons and axes of a late embryo.

A "LEC1 polypeptide" is a sequence of about 50 to about 210, sometimes 100 to 150, amino acid residues encoded by a *LEC1* polynucleotide. A full length LEC1 polypeptide and fragments containing a CCAAT binding factor (CBF) domain can act as a subunit of a protein capable of acting as a transcription factor in plant cells. LEC1 polypeptides are often distinguished by the presence of a sequence which is required for binding the nucleotide sequence: CCAAT. In particular, a short region of seven residues (MPIANVI) at residues 34-40 of SEQ. ID No. 3 shows a high degree of similarity to a region that has been shown to required for binding the CCAAT box. Similarly, residues 61-72 of SEQ. ID No. 3 (IQECVSEYISFV) is nearly identical to a region that contains a subunit interaction domain (Xing, *et al.*, (1993) *EMBO J.* 12:4647-4655).

As used herein, a homolog of a particular embryo-specific gene (*e.g.*, SEQ. ID. No. 1) is a second gene in the same plant type or in a different plant type, which has a polynucleotide sequence of at least 50 contiguous nucleotides which are substantially identical (determined as described below) to a sequence in the first gene. It is believed that, in general, homologs share a common evolutionary past.

A "polynucleotide sequence from" a particular embryo-specific gene is a subsequence or full length polynucleotide sequence of an embryo-specific gene which, when present in a transgenic plant, has the desired effect, for example, inhibiting expression of the endogenous gene driving expression of an heterologous polynucleotide. A full length sequence of a particular gene disclosed here may contain about 95%, usually at least about 98% of an entire sequence shown in the Sequence Listing, below.

In the case of both expression of transgenes and inhibition of endogenous genes (*e.g.*, by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical and may be "substantially identical" to a sequence of the gene from which it was derived. As explained below, these variants are specifically covered by this term.



In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the term "polynucleotide sequence from" a particular embryo-specific gene, such as *LECI*. In addition, the term specifically includes sequences (*e.g.*, full length sequences) substantially identical (determined as described below) with a *LECI* gene sequence and that encode proteins that retain the function of a *LECI* polypeptide.

In the case of polynucleotides used to inhibit expression of an endogenous gene, the introduced sequence need not be perfectly identical to a sequence of the target endogenous gene. The introduced polynucleotide sequence will typically be at least substantially identical (as determined below) to the target endogenous sequence.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The term "complementary to" is used herein to mean that the sequence is complementary to all or a portion of a reference polynucleotide sequence.

Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Add. APL. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (U.S.A.)* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

"Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched

positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 80% sequence identity, preferably at least 85%, more preferably at least 90% and most preferably at least 95%, compared to a reference sequence using the programs described above (preferably BLAST) using standard parameters. One of skill will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 40%, preferably at least 60%, more preferably at least 90%, and most preferably at least 95%. Polypeptides which are "substantially similar" share sequences as noted above except that residue positions which are not identical may differ by conservative amino acid changes. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, aspartic acid-glutamic acid, and asparagine-glutamine.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other, or a third nucleic acid, under stringent conditions. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5° C lower than the thermal melting point (T<sub>m</sub>) for the specific sequence at a defined ionic strength and pH. The T<sub>m</sub> is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least about 60°C.

In the present invention, mRNA encoded by embryo-specific genes of the invention can be identified in Northern blots under stringent conditions using cDNAs of the invention or fragments of at least about 100 nucleotides. For the purposes of this disclosure, stringent conditions for such RNA-DNA hybridizations are those which include at least one wash in 0.2X SSC at 63°C for 20 minutes, or equivalent conditions. Genomic DNA or cDNA comprising genes of the invention can be identified using the same cDNAs (or fragments of at least about 100 nucleotides) under stringent conditions, which for purposes of this disclosure, include at least one wash (usually 2) in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C, for 20 minutes, or equivalent conditions.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a restriction map of the 7.4 kb genomic wild-type fragment shown in SEQ. ID. No. 4.

#### DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention provides new embryo-specific genes useful in genetically engineering plants. Polynucleotide sequences from the genes of the invention can be used, for instance, to direct expression of desired heterologous genes in embryos (in the case of promoter sequences) or to modulate development of embryos or other organs (*e.g.*, by enhancing expression of the gene in a transgenic plant). In particular, the invention provides a new gene from *Arabidopsis* referred to here as *LEC1*. *LEC1* encodes polypeptides which subunits of a protein which acts as a transcription factor. Thus, modulation of the expression of this gene can be used to manipulate a number of useful traits, such as increasing or decreasing storage protein content in cotyledons or leaves.

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, 2nd. ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989).

Isolation of nucleic acids of the invention

The isolation of sequences from the genes of the invention may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here can be used to identify the desired gene in a cDNA or genomic DNA library from a desired plant species. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a library of embryo-specific cDNAs, mRNA is isolated from embryos and a cDNA library which contains the gene transcripts is prepared from the mRNA.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned embryo-specific gene such as the polynucleotides disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology to amplify the sequences of the genes directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes.

Appropriate primers and probes for identifying embryo-specific genes from plant tissues are generated from comparisons of the sequences provided herein. For a general overview of PCR see *PCR Protocols: A Guide to Methods and Applications*. (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), *Academic Press*, San Diego (1990). Appropriate primers for this purpose include, for instance: UP primer - 5' GGA ATT CAG CAA CAA CCC AAC CCC A 3' and LP primer - 5' GCT CTA GAC ATA CAA CAC TTT TCC TTA 3'. Alternatively, the following primer pairs can be used: 5' ATG ACC AGC TCA GTC ATA GTA GC 3' and 5' GCC ACA CAT GGT GGT TGC TGC TG 3' or 5' GAG ATA GAG ACC GAT CGT GGT TC 3' and 5' TCA CTT ATA CTG ACC ATA ATG GTC 3'. The amplifications conditions are typically as follows. Reaction components: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium

chloride, 0.001% gelatin, 200 microM dATP, 200 microM dCTP, 200 microM dGTP, 200 microM dTTP, 0.4 microM primers, and 100 units per ml Taq polymerase. Program: 96 C for 3 min., 30 cycles of 96 C for 45 sec., 50 C for 60 sec., 72 C for 60 sec, followed by 72 C for 5 min.

5 Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, e.g., Carruthers *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams *et al.*, *J. Am. Chem. Soc.* 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by  
10 adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Use of nucleic acids of the invention to inhibit gene expression

The isolated sequences prepared as described herein, can be used to prepare expression cassettes useful in a number of techniques. For example, expression cassettes of  
15 the invention can be used to suppress endogenous *LECI* gene expression. Inhibiting expression can be useful, for instance, in weed control (by transferring an inhibitory sequence to a weedy species and allowing it to be transmitted through sexual crosses) or to produce fruit with small and non-viable seed.

A number of methods can be used to inhibit gene expression in plants. For  
20 instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The expression cassette is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense RNA inhibits gene expression by preventing the accumulation of mRNA which  
25 encodes the enzyme of interest, see, e.g., Sheehy *et al.*, *Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt *et al.*, U.S. Patent No. 4,801,340.

The nucleic acid segment to be introduced generally will be substantially identical to at least a portion of the endogenous embryo-specific gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The  
30 vectors of the present invention can be designed such that the inhibitory effect applies to other proteins within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of at least about 500 nucleotides is especially preferred.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of embryo-specific genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff et al. *Nature*, 334:585-591 (1988).

Another method of suppression is sense suppression. Introduction of expression cassettes in which a nucleic acid is configured in the sense orientation with respect to the promoter has been shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli et al., *The Plant Cell* 2:279-289 (1990), and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184.

Generally, where inhibition of expression is desired, some transcription of the introduced sequence occurs. The effect may occur where the introduced sequence contains no coding sequence *per se*, but only intron or untranslated sequences homologous to

sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous sequences.

5 Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

For sense suppression, the introduced sequence in the expression cassette,  
10 needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants which are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-  
15 coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used.

Another means of inhibiting LEC1 function in a plant is by creation of dominant negatives. In this approach, non-functional, mutant LEC1 polypeptides, which retain the ability to interact with wild-type subunits are introduced into a plant. Identification  
20 of residues that can be changed to create a dominant negative can be determined by published work examining interaction of different subunits of CBF homologs from different species (*see, e.g., Sinha et al., (1995). Proc. Natl. Acad. Sci. USA 92:1624-1628.*)

#### Use of nucleic acids of the invention to enhance gene expression

25 Isolated sequences prepared as described herein can also be used to prepare expression cassettes which enhance or increase endogenous *LEC1* gene expression. Where overexpression of a gene is desired, the desired gene from a different species may be used to decrease potential sense suppression effects. Enhanced expression of *LEC1* polynucleotides is useful, for example, to increase storage protein content in plant tissues. Such techniques  
30 may be particularly useful for improving the nutritional value of plant tissues.

One of skill will recognize that the polypeptides encoded by the genes of the invention, like other proteins, have different domains which perform different functions.

Thus, the gene sequences need not be full length, so long as the desired functional domain of the protein is expressed. As explained above, LEC1 polypeptides share sequences with CBF proteins. The DNA binding activity, and, therefore, transcription activation function, of LEC1 polypeptides is thought to be modulated by a short region of seven residues

5 (MPIANVI) at residues 34-40 of SEQ. ID No. 2. Thus, the polypeptides of the invention will often retain these sequences. Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described for instance, in Sambrook *et al.*, *supra*. Hydroxylamine can also be used to introduce single base mutations into the coding region of the gene (Sikorski, *et al.*, (1991). *Meth. Enzymol.*  
10 194: 302-318). For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

Desired modified LEC1 polypeptides can be identified using assays to  
15 screen for the presence or absence of wild type LEC1 activity. Such assays can be based on the ability of the LEC1 protein to functionally complement the hap3 mutation in yeast. As noted above, it has been shown that homologs from different species functionally interact with yeast subunits of the CBF. (Sinha, *et al.*, (1995). *Proc. Natl. Acad. Sci. USA* 92:1624-1628); *see, also*, Becker, *et al.*, (1991). *Proc. Natl. Acad. Sci. USA* 88: 1968-1972). The reporter  
20 for this screen can be any of a number of standard reporter genes such as the lacZ gene encoding  $\beta$ -galactosidase that is fused with the regulatory DNA sequences and promoter of the yeast CYC1 gene. This promoter is regulated by the yeast CBF.

A plasmid containing the LEC1 cDNA clone is mutagenized *in vitro* according to techniques well known in the art. The cDNA inserts are excised from the plasmid and  
25 inserted into the cloning site of a yeast expression vector such as pYES2 (Invitrogen). The plasmid is introduced into hap3- yeast containing a lacZ reporter that is regulated by the yeast CBF such as pLG265UP1-lacZ (Guarente, *et al.*, (1984) *Cell* 36: 317-321). Transformants are then selected and a filter assay is used to test colonies for  $\beta$ -galactosidase activity. After confirming the results of activity assays, immunochemical tests using a LEC1 antibody are  
30 performed on yeast lines that lack  $\beta$ -galactosidase activity to identify those that produce stable LEC1 protein but lack activity. The mutant LEC1 genes are then cloned from the yeast and their nucleotide sequence determined to identify the nature of the lesions.



In other embodiments, the promoters derived from the *LEC1* genes of the invention can be used to drive expression of heterologous genes in an embryo-specific or seed-specific manner, such that desired gene products are present in the embryo, seed, or fruit. Suitable structural genes that could be used for this purpose include genes encoding proteins useful in increasing the nutritional value of seed or fruit. Examples include genes encoding enzymes involved in the biosynthesis of antioxidants such as vitamin A, vitamin C, vitamin E and melatonin. Other suitable genes encoding proteins involved in modification of fatty acids, or in the biosynthesis of lipids, proteins, and carbohydrates. Still other genes can be those encoding proteins involved in auxin and auxin analog biosynthesis for increasing fruit size, genes encoding pharmaceutically useful compounds, and genes encoding plant resistance products to combat fungal or other infections of the seed.

Typically, desired promoters are identified by analyzing the 5' sequences of a genomic clone corresponding to the embryo-specific genes described here. Sequences characteristic of promoter sequences can be used to identify the promoter. Sequences controlling eukaryotic gene expression have been extensively studied. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In most instances the TATA box is required for accurate transcription initiation. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. J. Messing et al., in *Genetic Engineering in Plants*, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)).

A number of methods are known to those of skill in the art for identifying and characterizing promoter regions in plant genomic DNA (see, e.g., Jordano, et al., *Plant Cell*, 1: 855-866 (1989); Bustos, et al., *Plant Cell*, 1:839-854 (1989); Green, et al., *EMBO J.* 7, 4035-4044 (1988); Meier, et al., *Plant Cell*, 3, 309-316 (1991); and Zhang, et al., *Plant Physiology* 110: 1069-1079 (1996)).

### 30 Preparation of recombinant vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for

transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, for example, Weising et al. *Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with  
5 transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for overexpression, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant.  
10 Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to  
15 those of skill.

Alternatively, the plant promoter may direct expression of the polynucleotide of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters). Examples of tissue-specific promoters under developmental control include promoters that initiate  
20 transcription only in certain tissues, such as fruit, seeds, or flowers. As noted above, the promoters from the *LECI* genes described here are particularly useful for directing gene expression so that a desired gene product is located in embryos or seeds. Other suitable promoters include those from genes encoding storage proteins or the lipid body membrane protein, oleosin. Examples of environmental conditions that may affect transcription by  
25 inducible promoters include anaerobic conditions, elevated temperature, or the presence of light.

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

30 The vector comprising the sequences (*e.g.*, promoters or coding regions) from genes of the invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide

resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

5     Production of transgenic plants

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA  
10     constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is  
15     infected by the bacteria.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *Embo J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm et al. *Proc. Natl. Acad. Sci. USA*  
20     82:5824 (1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:70-73 (1987).

*Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch et al. *Science* 233:496-498 (1984), and Fraley et al. *Proc. Natl. Acad. Sci. USA* 80:4803 (1983).  
25

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype such as seedlessness. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture  
30     growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture*,

*Handbook of Plant Cell Culture*, pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from the genera *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Cucumis*, *Cucurbita*, *Daucus*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Oryza*, *Panieum*, *Pannesetum*, *Persea*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Trigonella*, *Triticum*, *Vitis*, *Vigna*, and, *Zea*. The *LEC1* genes of the invention are particularly useful in the production of transgenic plants in the genus *Brassica*. Examples include broccoli, cauliflower, brussel sprouts, canola, and the like.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

#### Example 1

This example describes the isolation and characterization of an exemplary *LEC1* gene.

#### Experimental Procedures

##### Plant Material

A *lec1-2* mutant was identified from a population of *Arabidopsis thaliana* ecotype Wassilewskija (Ws-O) lines mutagenized with T-DNA insertions as described before (West et al., 1994). The *abi3-3*, *fus3-3* and *lec1-1* mutants were generously provided by Peter McCourt, University of Toronto and David Meinke, Oklahoma State University. Wild type plants and mutants were grown under constant light at 22°C.

Double mutants were constructed by intercrossing the mutant lines *lec1-1*, *lec1-2*, *abi3-3*, *fus3-3*, and *lec2*. The genotype of the double mutants was verified through

backcrosses with each parental line. Double mutants were those who failed to complement both parent lines. Homozygous single and double mutants were generated by germinating intact seeds or dissected mature embryos before desiccation on basal media.

#### Isolation and Sequence analysis of Genomic and cDNA Clones

5                    Genomic libraries of Ws-O wild type plants, *lec1-1* and *lec1-2* mutants were made in GEM11 vector according to the instructions of the manufacturer (Promega). Two silique-specific cDNA libraries (stages globular to heart and heart to young torpedo) were made in ZAPII vector (Stratagene).

10                   The genomic library of *lec1-2* was screened using right and left T-DNA specific probes according to standard techniques. About 12 clones that cosegregate with the mutation, were isolated and purified and the entire DNAs were further labeled and used as probes to screen a southern blot containing wild type and *lec1-1* genomic DNA. One clone hybridized with plant DNA and was further analyzed. A 7.1 kb XhoI fragment containing the left border and the plant sequence flanking the T-DNA was subcloned into  
15                   pBluescript-KS plasmid (Stratagene) to form ML7 and sequenced using a left border specific primer (5' GCATAGATGCACTCGAAATCAGCC 3'). The T-DNA organization was partially verified using southern analysis with T-DNA left and right borders and PBR322 probes. The results suggested that the other end of the T-DNA is also composed of left border. This was confirmed by generating a PCR fragment using a genomic plant DNA  
20                   primer (LP primer 5' GCT CTA GAC ATA CAA CAC TTT TCC TTA 3') and a T-DNA left border specific primer (5' GCTTGGTAATAATTGTCATTAG 3') and sequencing.

                    The EcoRI insert of ML7 was used to screen a wild type genomic library. Two overlapping clones were purified and a 7.4 EcoRI genomic fragment from the wild type DNA region was subcloned into pBluescript-KS plasmid making WT74. This fragment was  
25                   sequenced (SEQ. ID. No. 4) and was used to screen *lec1-1* genomic library and wild type silique-specific cDNA libraries. 8 clones from the *lec1-1* genomic library were identified and analyzed by restriction mapping.

                    From these clones the exact site of the deletion in *lec1-1* was mapped and sequenced by amplifying a Xbp PCR fragment using primers (H21 - 5' H21 - 5' CTA AAA  
30                   ACA TCT ACG GTT CA 3'; H 17 - 5' TTT GTG GTT GAC CGT TTG GC 3') flanking the deletion region in *lec1-1* genomic DNA. Clones were isolated from both cDNA libraries and partially sequenced. The sequence of the cDNA clones and the wild type genomic clone

matched exactly, confirming that both derived from the same locus. All hybridizations were performed under stringent conditions with 32P random prime probes (Stratagene).

Sequencing was done using the automated dideoxy chain termination method (Applied Biosystems, Foster City, CA). Data base searches were performed at the National Center for Biotechnology Information by using the BLAST network service. Alignment of protein sequences was done using PILEUP program (Genetics Computer Group, Madison, WI)

#### DNA and RNA blot analysis

Genomic DNA was isolated from leaves by using the CTAB-containing buffer Dellaporta, *et al.*, (1983). *Plant Mol. Biol. Reporter* 1: 19-21. Two micrograms of DNA was digested with different restriction endonucleases, electrophoretically separated in 1% agarose gel, and transferred to a nylon membrane (Hybond N; Amersham).

Total RNA was prepared from siliques, two days old seedlings, stems, leaves, buds and roots. Poly(A)+ RNA was purified from total RNA by oligo(dT) cellulose chromatography, and two micrograms of each Poly(A)+ RNA samples were separated in 1% denatured formaldehyde-agarose gel. Hybridizations were done under stringent conditions unless it specifies otherwise. Radioactive probes were prepared as described above.

#### Complementation of lec1 mutants

A 3.4 kb BstI fragment of genomic DNA (SEQ. ID. No. 3) containing sequences from 1.992 kb upstream of the ORF to a region 579 bp downstream from the poly A site was subcloned into the hygromycin resistant binary vector pBIB-Hyg. The LEC1 cDNA was placed under the control of the 35S promoter and the ocs polyadenylation signals by inserting a PCR fragment spanning the entire coding region into the plasmid pART7. The entire regulatory fragment was then removed by digestion with NotI and transferred into the hygromycin resistant binary vector BJ49. The binary vectors were introduced into the *Agrobacterium* strain GV3101, and constructions were checked by re-isolation of the plasmids and restriction enzyme mapping, or by PCR. Transformation to homozygous lec1-1 and lec1-2 mutants were done using the in planta transformation procedure (Bechtold, *et al.*, (1993). *Comptes Rendus de l'Academie des Sciences Serie III Sciences de la Vie*, 316: 1194-1199. Dry seeds from lec1 mutants were selected for transformants by their ability to germinate after desiccation on plates containing 5g/ml hygromycin. The transformed plants

were tested for the presence of the transgene by PCR and by screening the siliques for the presence of viable seeds.

#### In Situ Hybridization

- Experiments were performed as described previously by Dietrich *et al.* (1989)
- 5 Plant Cell 1: 73-80. Sections were hybridized with LEC1 antisense probe. As a negative control, the LEC1 antisense probe was hybridized to seed sections of *lec1* mutants. In addition, a sense probe was prepared and reacted with the wild type seed sections.

#### Results

##### 10 Genetic Interaction Between Leafy Cotyledon-Type Mutants and *abi3*

- In order to understand the genetic pathways which regulate late embryogenesis we took advantage of three Arabidopsis mutants *lec2*, *fus3-3* and *abi3-3* that cause similar defects in late embryogenesis to those of *lec1-1* or *lec1-2*. These mutants are desiccation intolerant, sometimes viviparous and have activated shoot apical meristems. The *lec2* and
- 15 *fus3-3* mutants are sensitive to ABA and possess trichomes on their cotyledons and therefore can be categorized as leafy cotyledon-type mutants (Meinke *et al.*, 1994). The *abi3-3* mutants belong to a different class of late embryo defective mutations that is insensitive to ABA and does not have trichomes on the cotyledons.

- The two classes of mutants were crossed to *lec1-1* and *lec1-2* mutants to
- 20 construct plants homozygous to both mutations. The *lec1* and *lec2* mutations interact synergistically, resulting in a double mutant which is arrested in a stage similar to the late heart stage, the double mutant embryo, however, is larger. The *lec1* or *lec2* and *fus3-3* double mutants did not display any epistasis and the resulting embryo had an intermediate phenotype. The *lec1/abi3-3* double mutants and *lec2/abi3-3* double mutants were ABA
- 25 insensitive and had a *lec*-like phenotype. There was no difference between double mutants that consist of either *lec1-1* or *lec1-2*.

No epistasis was seen between the double mutants indicating that each of the above genes, the LEC-type and ABI3 genes, operate in different genetic pathways.

#### LEC1 Functions Early in Embryogenesis

- 30 The effects of *lec1* is not limited to late embryogenesis, it also has a role in early embryogenesis. The embryos of the *lec1/lec2* double mutants were arrested in the early

stages of development, while the single mutants developed into mature embryos, suggesting that these genes act early during development.

Further examination of the early stages of the single and double mutations showed defects in the shape, size and cell division pattern of the mutants suspensors. The suspensor of wild type embryo consists of a single file of six to eight cells, whereas the suspensors of the mutants are often enlarged and undergo periclinal divisions. Leafy cotyledon mutants exhibit suspensor anomalies at the globular or transition stage whereas wild type and *abi3* mutant do not show any abnormalities.

The number of anomalous suspensors increases as the embryos continue to develop. At the torpedo stage, the wild type suspensor cells undergo programmed cell death, but in the mutants secondary embryos often develop from the abnormal suspensors and, when rescued, give rise to twins.

#### The Organization of the *LEC1* Locus in Wild Type Plants and *lec1* Mutants

Two mutant alleles of the *LEC1* gene have been reported, *lec1-1* and *lec1-2* (Meinke, 1992; West et al., 1994). Both mutants were derived from a population of plants mutagenized insertionally with T-DNA (Feldmann and Marks, 1987), although *lec1-1* is not tagged. The *lec1-2* mutant contains multiple T-DNA insertions. A specific subset of T-DNA fragments were found to be closely linked with the mutation. A genomic library of *lec1-2* was screened using right and left borders T-DNA as probes. Genomic clones containing T-DNA fragments that cosegregate with the mutation were isolated and tested on southern blots of both wild type and *lec1-1* plants. Only one clone hybridized with *Arabidopsis* DNA and also gave polymorphic restriction fragment in *lec1-1*.

The *lec1-1* polymorphism resulted from a small deletion, approximately 2 kb in length. Using sequences from the plant fragment flanking the T-DNA, the genomic wild type DNA clones and the *lec1-1* genomic clones were isolated. An *EcoRI* fragment of 7.4 kb of the genomic wild type DNA that corresponded to the polymorphic restriction fragment in *lec1-1* was further analyzed and sequenced. The exact site of the deletion in *lec1-1* was identified using a PCR fragment that was generated by primers, within the expected borders of the deleted fragment, and sequencing.

In the wild type genomic DNA that corresponded to the *lec1-1* deletion, a 626 bp ORF was identified. Southern analysis of wild type DNA and the two mutants DNA probed with the short DNA fragment of the ORF revealed that both the wild type and *lec1-2*



DNA contain the ORF while the *lec1-1* genomic DNA did not hybridize. The exact insertion site of the T-DNA in *lec1-2* mutant was determined by PCR and sequencing and it was found that the T-DNA was inserted 115 bp upstream of the ORF's translational initiation codon in the 5' region of the gene.

5                   At the site of the T-DNA insertion a small deletion of 21 plant nucleic acids and addition of 20 unknown nucleic acids occurred. These results suggest that in *lec1-2* the T-DNA interferes with the regulation of the ORF while in *lec1-1* the whole gene is deleted. Thus, both *lec1* alleles contain DNA disruptions at the same locus, confirming the identity of the *LEC1* locus.

10           The *lec1* Mutants Can Be Complemented by Transformation

                  To prove that the 7.4 kb genomic wild type fragment indeed contained the ORF of the *LEC1* gene, we used a genomic fragment of 3395 bp (SEQ. ID. No. 3) within that fragment to transform homozygous *lec1-1* and *lec1-2* plants. The clone consists of a 3395 bp BstYI restriction fragment containing the gene and the promoter region. The  
15 translation start codon (ATG) of the polypeptide is at 1999 and the stop codon is at 2625 (TGA). There are no introns in the gene.

                  The transformed plants were selected on hygromycin plates and were tested to contain the wild type DNA fragment by PCR analysis. Both transgenic mutants were able to produce viable progeny, that were desiccation tolerant and did not possess trichomes on their  
20 cotyledons. We concluded that the 3.4 kb fragment can complement the *lec1* mutation and since there is only one ORF in the deleted 2 kb fragment in *lec1-1* we suggest that this ORF corresponds to the *LEC1* gene.

The *LEC1* Gene is a Member of Gene Family

                  In order to isolate the *LEC1* gene two cDNA libraries of young siliques were  
25 screened using the 7.4 kb DNA fragment as a probe. Seventeen clones were isolated and after further analysis and partial sequencing they were all found to be identical to the genomic ORF. The cDNA contains 626 bp ORF specifying 208 amino acid protein (SEQ. ID. Nos. 1 and 2).

                  The *LEC1* cDNA was used to hybridize a DNA gel blot containing Ws-O  
30 genomic DNA digested with three different restriction enzymes. Using low stringency hybridization we found that there is at least one more gene. This confirmed our finding of two more Arabidopsis ESTs that show homology to the *LEC1* gene.

The *LEC1* gene is Embryo Specific

The *lec1* mutants are affected mostly during embryogenesis. Rescued mutants can give rise to homozygous plants that have no obvious abnormalities other than the presence of trichomes on their cotyledons and their production of defective progeny.

- 5 Therefore, we expected the *LEC1* gene to have a role mainly during embryogenesis and not during vegetative growth. To test this assumption Poly (A)+ RNA was isolated from siliques, seedling, roots, leaves, stems and buds of wild type plants and from siliques of *lec1* plants. Only one band was detected on northern blots using either the *LEC1* gene as a probe or the 7.4 kb genomic DNA fragment suggesting that there is only one gene in the genomic DNA
- 10 fragment which is active transcriptionally. The transcript was detected only in siliques containing young and mature embryos and was not detected in seedlings, roots, leaves, stems and buds indicating that the *LEC1* gene is indeed embryo specific. In addition, no RNA was detected in siliques of both alleles of *lec1* mutants confirming that this ORF corresponds to the *LEC1* gene.

15 Expression Pattern of the *LEC1* Gene

To study how the *LEC1* gene specifies cotyledons identity, we analyzed its expression by in situ hybridization. We specifically focused on young developing embryos since the mutants abnormal suspensors phenotype indicates that the *LEC1* gene should be active very early during development.

- 20 During embryogenesis, the *LEC1* transcript was first detected in proglobular embryos. The transcript was found in all cells of the proembryo and was also found in the suspensor and the endosperm. However, from the globular stage and on it accumulates more in the outer layer of the embryo, namely the protoderm and in the outer part of the ground meristem leaving the procambium without a signal. At the torpedo stage the signal was
- 25 stronger in the cotyledons and the root meristem, and was more limited to the protoderm layer. At the bent cotyledon stage the signal was present throughout the embryo and at the last stage of development when the embryo is mature and filling the whole seed we could not detect the *LEC1* transcript. This might be due to sensitivity limitation and may imply that if the *LEC1* transcript is expressed at that stage it is not localized in the mature embryo, but
- 30 rather spread throughout the embryo.

The *LEC1* gene encodes a Homolog of CCAAT binding factor.

Comparison of the deduced amino acid sequence of LEC1 to the GeneBank reveals significant similarity to a subunit of a transcription factor, the CCAAT box binding factor (CBF). CBFs are highly conserved family of transcription factors that regulate gene activity in eukaryotic organisms Mantvani, *et al.*, . (1992). *Nucl. Acids Res.* 20: 1087-1091.

5 They are hetero-oligomeric proteins that consist of between three to four non-homologous subunits. LEC1 was found to have high similarity to CBF-A subunit. This subunit has three domains; A and C which show no conservation between kingdoms and a central domain, B, which is highly conserved evolutionary. Similarly the *LEC1* gene is composed of three domains. The *LEC1* B domain shares between 75%-85% similarity and 55%-63% identity

10 with different B domains that are found in organisms ranging from yeast to human. Within this central domain, two highly conserved amino acid segments are present. Deletion and mutagenesis analysis in the CBF-A yeast homolog hap3 protein demonstrated that a short region of seven residues (42-48) (LPIANVA) is required for binding the CCAAT box, while the subunit interaction domain lies in the region between residues 69-80 (MQECVSEFISFV)

15 (Xing *et al.*, *supra*). LEC1 protein shares high homology to those regions.

## **DISCUSSION**

The *lec1* mutant belongs to the leafy cotyledon class that interferes mainly with the embryo program and therefore is thought to play a central regulatory role during

20 embryo development. It was shown before that *LEC1* gene activity is required to suppress germination during the maturation stage. Therefore, we analyzed the genetic interaction of homozygous double mutants of the different members of the leafy cotyledon class and the *abi3* mutant that has an important role during embryo maturation. All the five different combinations of the double mutants showed either an intermediate phenotype or an additive

25 effect. No epistatic relationship among the four genes was found. These findings suggest that the different genes act in parallel genetic pathways. Of special interest was the double mutant *lec1/lec2* that was arrested morphologically at the heart stage, but continued to grow in that shape. This double mutant phenotype indicates that both genes *LEC1* and *LEC2* are essential for early morphogenesis and their products may interact directly or indirectly in the

30 young developing embryo.

### **The Role of *LEC1* in Embryogenesis**

One of the proteins that mediate CCAAT box function, is an heteromeric protein called CBF (also called NFY or CP1). CBF is a transcription activator that regulates constitutively expressed genes, but also participates in differential activation of developmental genes Wingender, E. (1993). *Gene Regulation in Eukaryotes* (New York: VCH Publishers).

5 In mammalian cells, three subunits have been identified CBF-A, CBF-B and CBF-C and all of which are required for DNA binding. In yeast, the CBF homolog HAP activates the CYC1 and other genes involved in the mitochondrial electron transport Johnson, *et al.*, *Proteins. Annu. Rev. Biochem.* 58, 799-840. (1989). HAP consists of four subunits hap2, hap3, hap4 and hap5. Only hap2, 3 and 5 are required for DNA binding. CBF-A, B and C show high  
10 similarity to the yeast hap3, 2 and 5, respectively. It was also reported that mammalian CBF-A and B can be functionally interchangeable with the corresponding yeast subunits (Sinha *et al.*, *supra.*).

The *LEC1* gene encodes a protein that shows more then 75% similarity to the conserved region of CBF-A. CCAAT motifs are not common in plants' promoters and their  
15 role in transcription regulation is not clear. However, maize and *Brassica* homologs have been identified Search in the Arabidopsis GeneBank revealed several ESTs that show high similarity to CBF-A, B and C. Accession numbers of CBF-A (HAP3) homologs: H37368, H76589; CBF-B (HAP2) homologs: T20769; CBF-C (HAP5) homologs: T43909, T44300. These findings and the pleiotropic affects of *LEC1* suggest that *LEC1* is a member of a  
20 heteromeric complex that functions as a transcription factor.

The model suggests that *LEC1* acts as transcription activator to several sets of genes, which keep the embryonic program on and repress the germination process. Defective *LEC1* expression partially shuts down the embryonic program and as a result the cotyledons lose their embryonic characteristics and the germination program is active in the  
25 embryo.

### Example 2

This example demonstrates that *LEC1* is sufficient to induce embryonic pathways in transgenic plants.

30 The phenotype of *lec1* mutants and the gene's expression pattern indicated that *LEC1* functions specifically during embryogenesis. A *LEC1* cDNA clone under the

control of the cauliflower mosaic virus 35S promoter was transferred into *lec1-1* mutant plants *in planta* using standard methods as described above.

Viable dry seeds were obtained from *lec1-1* mutants transformed with the 35S/*LEC1* construct. However, the transformation efficiency was only approximately 0.6% of that obtained normally. In several experiments, half the seeds that germinated (12/23) produced seedlings with an abnormal morphology. Unlike wild type seedlings, these 35S/*LEC1* seedlings possessed cotyledons that remained fleshy and that failed to expand. Roots often did not extend or extended abnormally and sometimes greened. These seedlings occasionally produced a single pair of organs on the shoot apex at the position normally occupied by leaves. Unlike wild type leaves, these organs did not expand and did not possess trichomes. Morphologically, these leaf-like structures more closely resembled embryonic cotyledons than leaves.

The other 35S/*LEC1* seeds that remained viable after drying produced plants that grow vegetatively. The majority of these plants (7) flowered and produced 100% *lec1* mutant seeds. Amplification experiments confirmed that the seedlings contained the transgene, suggesting that the 35S/*LEC1* gene was inactive in these T2 seeds. No vegetative abnormalities were observed in these plants with the exception that a few displayed defects in apical dominance. A few plants (2) were male sterile and did not produce progeny. One plant that produced progeny segregated 25% mutant *Lec1*<sup>-</sup> seeds that, when germinated before desiccation and grown to maturity, gave rise to 100% mutant seed, as expected for a single transgene locus. The other 75% of seeds contained embryos with either a wild type phenotype or a phenotype intermediate between *lec1* mutants and wild type. Only 25% of the dry seed from this plant germinated, and all seedlings resembled the embryo-like seedlings described above. Some seedlings continued to grow and displayed a striking phenotype. These 35S/*LEC1* plants developed two types of structures on leaves. One type resembled embryonic cotyledons while the other looked like intact torpedo stage embryos. Thus, ectopic expression of *LEC1* induces the morphogenesis phase of embryo development in vegetative cells.

Because many 35S/*LEC1* seedlings exhibited embryonic characteristics, the seedlings were analyzed for expression of genes specifically active in embryos. Cruciferin A storage protein mRNA accumulated throughout the 35S/*LEC1* seedlings, including the leaf-like structures. Proteins with sizes characteristic of 12S storage protein cruciferin

accumulated in these transgenic seedlings. Thus, *35S/LEC1* seedlings displaying an embryo-like phenotype accumulated embryo-specific mRNAs and proteins. *LEC1* mRNA accumulated to a high level in these *35S/LEC1* seedlings in a pattern similar to early stage embryos but not in wild type seedlings. *LEC1* is therefore sufficient to alter the fate of vegetative cells by inducing embryonic programs of development.

The ability of *LEC1* to induce embryonic programs of development in vegetative cells establishes the gene as a central regulator of embryogenesis. *LEC1* is sufficient to induce both the seed maturation pathway as indicated by the induction of storage protein genes in the *35S/LEC1* seedlings. The presence of ectopic embryos on leaf surfaces and cotyledons at the position of leaves also shows that *LEC1* can activate the embryo morphogenesis pathway. Thus, *LEC1* regulates both early and late embryonic processes.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: The Regents of the University of California
- (ii) TITLE OF INVENTION: Leafy Cotyledon1 Genes and Their Uses
- 5 (iii) NUMBER OF SEQUENCES: 18
- (iv) CORRESPONDENCE ADDRESS:
- 10 (A) ADDRESSEE: Fulbright & Jaworski, LLP
- (B) STREET: 865 S. Figueroa Street, 29th Floor
- (C) CITY: Los Angeles
- (D) STATE: California
- (E) COUNTRY: USA
- (F) ZIP: 90017-2571
- 15 (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- 20 (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
- 25 (A) NAME: Berliner, Robert
- (B) REGISTRATION NUMBER: 20,121
- (C) REFERENCE/DOCKET NUMBER: 5555-470
- (ix) TELECOMMUNICATION INFORMATION:
- 30 (A) TELEPHONE: (213) 892-9200
- (B) TELEFAX: (213) 680-4518
- (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:
- 35 (A) LENGTH: 627 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- 40 (ix) FEATURE:
- (A) NAME/KEY: CDS
- (B) LOCATION: 1..627

## (D) OTHER INFORMATION: /product= "LEC1"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

5	ATG ACC AGC TCA GTC ATA GTA GCC GGC GCC GGT GAC AAG AAC AAT GGT	48
	Met Thr Ser Ser Val Ile Val Ala Gly Ala Gly Asp Lys Asn Asn Gly	
	1 5 10 15	
10	ATC GTG GTC CAG CAG CAA CCA CCA TGT GTG GCT CGT GAG CAA GAC CAA	96
	Ile Val Val Gln Gln Gln Pro Pro Cys Val Ala Arg Glu Gln Asp Gln	
	20 25 30	
15	TAC ATG CCA ATC GCA AAC GTC ATA AGA ATC ATG CGT AAA ACC TTA CCG	144
	Tyr Met Pro Ile Ala Asn Val Ile Arg Ile Met Arg Lys Thr Leu Pro	
	35 40 45	
20	TCT CAC GCC AAA ATC TCT GAC GAC GCC AAA GAA ACG ATT CAA GAA TGT	192
	Ser His Ala Lys Ile Ser Asp Asp Ala Lys Glu Thr Ile Gln Glu Cys	
	50 55 60	
25	GTC TCC GAG TAC ATC AGC TTC GTG ACC GGT GAA GCC AAC GAG CGT TGC	240
	Val Ser Glu Tyr Ile Ser Phe Val Thr Gly Glu Ala Asn Glu Arg Cys	
	65 70 75 80	
30	CAA CGT GAG CAA CGT AAG ACC ATA ACT GCT GAA GAT ATC CTT TGG GCT	288
	Gln Arg Glu Gln Arg Lys Thr Ile Thr Ala Glu Asp Ile Leu Trp Ala	
	85 90 95	
35	ATG AGC AAG CTT GGG TTC GAT AAC TAC GTG GAC CCC CTC ACC GTG TTC	336
	Met Ser Lys Leu Gly Phe Asp Asn Tyr Val Asp Pro Leu Thr Val Phe	
	100 105 110	
40	ATT AAC CGG TAC CGT GAG ATA GAG ACC GAT CGT GGT TCT GCA CTT AGA	384
	Ile Asn Arg Tyr Arg Glu Ile Glu Thr Asp Arg Gly Ser Ala Leu Arg	
	115 120 125	
45	GGT GAG CCA CCG TCG TTG AGA CAA ACC TAT GGA GGA AAT GGT ATT GGG	432
	Gly Glu Pro Pro Ser Leu Arg Gln Thr Tyr Gly Gly Asn Gly Ile Gly	
	130 135 140	
50	TTT CAC GGC CCA TCT CAT GGC CTA CCT CCT CCG GGT CCT TAT GGT TAT	480
	Phe His Gly Pro Ser His Gly Leu Pro Pro Pro Gly Pro Tyr Gly Tyr	
	145 150 155 160	
55	GGT ATG TTG GAC CAA TCC ATG GTT ATG GGA GGT GGT CGG TAC TAC CAA	528
	Gly Met Leu Asp Gln Ser Met Val Met Gly Gly Gly Arg Tyr Tyr Gln	
	165 170 175	
60	AAC GGG TCG TCG GGT CAA GAT GAA TCC AGT GTT GGT GGT GGC TCT TCG	576
	Asn Gly Ser Ser Gly Gln Asp Glu Ser Ser Val Gly Gly Gly Ser Ser	
	180 185 190	
65	TCT TCC ATT AAC GGA ATG CCG GCT TTT GAC CAT TAT GGT CAG TAT AAG	624
	Ser Ser Ile Asn Gly Met Pro Ala Phe Asp His Tyr Gly Gln Tyr Lys	
	195 200 205	
	TGA	627



## (2) INFORMATION FOR SEQ ID NO:2:

5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 208 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr Ser Ser Val Ile Val Ala Gly Ala Gly Asp Lys Asn Asn Gly  
 1 5 10 15  
 15 Ile Val Val Gln Gln Gln Pro Pro Cys Val Ala Arg Glu Gln Asp Gln  
 20 25 30  
 Tyr Met Pro Ile Ala Asn Val Ile Arg Ile Met Arg Lys Thr Leu Pro  
 20 35 40 45  
 Ser His Ala Lys Ile Ser Asp Asp Ala Lys Glu Thr Ile Gln Glu Cys  
 50 55 60  
 25 Val Ser Glu Tyr Ile Ser Phe Val Thr Gly Glu Ala Asn Glu Arg Cys  
 65 70 75 80  
 Gln Arg Glu Gln Arg Lys Thr Ile Thr Ala Glu Asp Ile Leu Trp Ala  
 85 90 95  
 30 Met Ser Lys Leu Gly Phe Asp Asn Tyr Val Asp Pro Leu Thr Val Phe  
 100 105 110  
 Ile Asn Arg Tyr Arg Glu Ile Glu Thr Asp Arg Gly Ser Ala Leu Arg  
 35 115 120 125  
 Gly Glu Pro Pro Ser Leu Arg Gln Thr Tyr Gly Gly Asn Gly Ile Gly  
 130 135 140  
 40 Phe His Gly Pro Ser His Gly Leu Pro Pro Pro Gly Pro Tyr Gly Tyr  
 145 150 155 160  
 Gly Met Leu Asp Gln Ser Met Val Met Gly Gly Gly Arg Tyr Tyr Gln  
 165 170 175  
 45 Asn Gly Ser Ser Gly Gln Asp Glu Ser Ser Val Gly Gly Gly Ser Ser  
 180 185 190  
 50 Ser Ser Ile Asn Gly Met Pro Ala Phe Asp His Tyr Gly Gln Tyr Lys  
 195 200 205

## (2) INFORMATION FOR SEQ ID NO:3:

55 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3395 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

	AGATCCAAAA CAGGTCATGG ACTGGGCCGT AAATCTATC CAAAATCTT CATGTTTTTC	60
10	CATCTTTCAA AAATCTTAT CCACCATTCC ATTACTAGG TGTTGGTTT ATTTTATTG	120
	TTGATTAATT ATGTATTAGA AAATGTAAAG CAATATTCAA TTGTAACATG CATCATCTAA	180
15	CACCAATATC TTGTACTAAC CTTTTGTAAT TTTCTATAA ACATTTTAA AGGCTAATTT	240
	AAATAAAAAT TACAATAAAC GTGATAACTC ACTTTCGTAA CGCATATTTA TTCAAATATA	300
	CCAAAATTTA CCATTTTAAG TAAGAGAATC TTTTAAAAAT TAATTTTCAA TTTCATTAAT	360
20	TAAGAAACAA AGAATTTACT GAAACCTATA TTTTATTAA TTTTAATAAA ATATATGACT	420
	AAAATAACGT CACGTGAATC TTTCTCAGCC GTTCGATAAT CGAATACTTT ATTGACTAAG	480
25	TATTTATTTA GAAAATTTTA AACAACACTT AATTTCTAGA AACAAAGAGA GCCTCATATG	540
	TATAAAAATC TTCTTCTTAT CTTTCTTCT TTCTTAATAG TCTTTATTTT TACTTAATTA	600
	CTTTGGTAAT TTGTGAAAA CACAACCAAT GAGAGAAGAG CAGTTTGACT GGCCACATAG	660
30	CCAATGAGAC AAGCCAATGG GAAAGAGATA TAGAGACCTC GTAAGAACCG CTCCTTTGCC	720
	ATTTGTATCA TCTCTCTATA AAACCACTCA ACCATCAACC TNTCTTTGCA TGCAACAAAT	780
35	CACTCAAATA ATTATTTTAT AAAGAACAAA AAAAAAAGA CGGCAGAGAA ACAATGGAAC	840
	GTGGAGCTCC CTTCTCTCAC TATCAGCTAC CCAAATCCAT CTCTGGTAAT CTAAGTGGCT	900
	ATTTGTATAC AGTATATACT TGCCTCCATG TATATTTATA TTCTCGTGAA AAATTGGAGA	960
40	CATGCTTTAT GAATTTTATG AGACTTTGCA ACAACGAACG AGATGCTTTC TCTCTAGAAA	1020
	TTTAAATTTA GATTTGTGAA GGTTTTGGGA ATGGCCCGGA GAAGACGATT TTATATATAC	1080
45	ATGCATGCAA GAGTTTGATA TGTATATTGT TTCATCATGG CTGAGTCAA GTTTTATCCA	1140
	AATATTTCCA TGGTGTGGTA TTAGTTAAAC AAATCTCTCG TATGTGTCAT TGAATATACC	1200
	CGTGCATGTA CCAGGAATGT TTTTGATTCT AAAAACGTTT TTTTCTTGT TGTAACGGTT	1260
50	GAGTTTTTTT CTCGTTTCA AAACGAGATT CTCGTTTGTC TCTCCCTTG TCTAAAAACA	1320
	TCTACGGTTC ATGTGATTCA AAAACACTAA AAAAATATAA ACTCATTTTT TTTAATACT	1380
55	TAACATTTAA ACTATATATA TATATATATA TATATATATC TTATACTAGT CCCAAGTTTT	1440
	AGTGTGAGGT TTTTATTATC AAAATCTATC AGTACATTTT TTGGAAGA ACTAAGTGAA	1500
	ATTTTCTCCA AATTTTCCTT TTAATTTGA TTTTAAATT ACTGGATGTC ATTAACTTTA	1560

	ATCTTTTGAT	TCTTTCAACG	TTTACCATTG	GGAACCTTCA	CATGAAATAA	ATGTCTACTT	1620
	TATTGAGTCA	TACCTTCGTC	AACATAAATT	AATTGATGTT	CTTCTCCAAA	TTTTGAGTTT	1680
5	TTGGTTTTTC	TAATAATCTT	AACGAAAGCT	TTTTGGTATA	CATGTAAAAC	GTAACGGCAA	1740
	GAATCTGAAC	AGTCTACTCA	ACGGGGTCCA	TAAGTCTAGA	ATGTAGACCC	CACAACTTA	1800
10	CTCTTATCTT	ATTGGTCCGT	AACTAAGAAC	GTGTCCCTCT	GATTCTCTTG	TTTTCTTCTA	1860
	ATTAATTCGT	ATCCTACAAA	TTTAATTATC	ATTTCTACTT	CAACTAATCT	TTTTTTATTT	1920
	CCTAAAGATT	TCAATTTCTC	TCTGTATTTT	CTATGAACAG	AATTGAACTT	GGACCAGCAC	1980
15	AGCAACAACC	CAACCCCAAT	GACCAGCTCA	GTCATAGTAG	CCGGCGCCGG	TGACAAGAAC	2040
	AATGGTATCG	TGGTCCAGCA	GCAACCACCA	TGTGTGGCTC	GTGAGCAAGA	CCAATACATG	2100
20	CCAATCGCAA	ACGTCATAAG	AATCATGCGT	AAAACCTTAC	CGTCTCACGC	CAAATCTCTT	2160
	GACGACGCCA	AAGAAACGAT	TCAAGAATGT	GTCTCCGAGT	ACATCAGCTT	CGTGACCGGT	2220
	GAAGCCAACG	AGCGTTGCCA	ACGTGAGCAA	CGTAAGACCA	TAAGTGCTGA	AGATATCCTT	2280
25	TGGGCTATGA	GCAAGCTTGG	GTTGATAAC	TACGTGGACC	CCCTCACCGT	GTTTCATTAAC	2340
	CGGTACCGTG	AGATAGAGAC	CGATCGTGGT	TCTGCACTTA	GAGGTGAGCC	ACCGTCGTTG	2400
30	AGACAAACCT	ATGGAGGAAA	TGGTATTGGG	TTTCACGGCC	CATCTCATGG	CCTACCTCCT	2460
	CCGGGTCCTT	ATGGTTATGG	TATGTTGGAC	CAATCCATGG	TTATGGGAGG	TGGTCGGTAC	2520
	TACCAAACG	GGTCGTCGGG	TCAAGATGAA	TCCAGTGTTG	GTGGTGGCTC	TTCGTCTTCC	2580
35	ATTAACGGAA	TGCCGGCTTT	TGACCATTAT	GGTCAGTATA	AGTGAAGAAG	GAGTTATTCT	2640
	TCATTTTTAT	ATCTATTCAA	AACATGTGTT	TCGATAGATA	TTTTATTTTT	ATGTCTTATC	2700
40	AATAACATTT	CTATATAATG	TTGCTTCTTT	AAGGAAAAGT	GTTGTATGTC	AATACTTTAT	2760
	GAGAAACTGA	TTTATATATG	CAAATGATTG	AATCCAAACT	GTTTTGTGGA	TTAAACTCTA	2820
	TGCAACATTA	TATATTTACA	TGATCTAAAG	GTTTTGTAAT	TCAAAAGCTG	TCATAGTTAG	2880
45	AAGATAACTA	AACATTGTAG	TAACCAAGTT	TAATTTACTT	TTTTGAGTTT	ACATAACTAA	2940
	CCAAGCCAAA	AGGTTATAAA	ATCTAAATTC	GTTGAGTTGT	CAAACTTCTG	AAGATTGCTA	3000
50	TCCTCTTTGA	GTTGCTTTCT	TTTGGGTGCT	TGAGTTTCAT	TAGGCTGAGC	TGACTCGTTG	3060
	CTCTCTAGTC	TTTCATCTCT	GTCTTTTCCA	AGGATTCATA	ACGTTGGTCG	CTCTCTGTTT	3120
	CTGCCCTACAC	TTCTTCAAGG	GATCATTACT	GAGGCTAAGA	GTTAAAGACC	TGAACCATGG	3180
55	TTTTCTGTAA	CTGGTTCAAG	TTCATTCTCC	GGTTATTGTG	TGGTTATCTT	TCGGTTAGAT	3240
	TGAAACCCAT	ATGTTTGCTC	TGTTTCTTCT	AGTTCCAAGT	TTAATTTCCG	GTTATTGTTT	3300

GGCTTTTAA AAGTTTTTAA GGTCTATTCT ATGTAAAGAC TATTCTACGT ACGTACATTT 3360  
ATCGCAAAAT TGAAAGATTA TAAAAAAAT TGAAA 3395

5

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 7560 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

20 AATTNACCCT CACTAAAGGG AACAAAAGCT GGGTACCGGG CCCCCCTCG AGGTCGACGG 60  
TATCGATAAG CTTGATATCG AATTCGTGGC CATTAGACCC ATAACCTATAT GACGATGTTA 120  
AAGAGAAAAT AAATCATAAA TAAATAAGA GTCCTTATCA ATAAACCTAA TTGGCTAATT 180  
25 TCAACCTCAA AGAGTAGTAG GAACAGGTAA GGTGAAGCCA AACAGCTCCT TTTACAGTTG 240  
GACCACTAGA GCTGATCTGG CATACAAAGT ATGCTTATTG GGCTGTCACG GCCCATCCGC 300  
AAAATGTCGT TGGTTACGAA GCATCCACGA CATAGACGGT GCCACATGTT AGAAAAGTGT 360  
30 TTCGGCGATC AAGATTGTGT CCACATCATT AGACGTCTGA ACTGTCCACG TGTCTATCAA 420  
AGCTGGCGTC AAACATTACG TTTTCGTCGT TTGCGCCTCC TAGTTCACAC GTGCAACGAA 480  
35 CGCGTGCGAC GTATCAAAAT TGTTAATTTT AGCCATGTAT AAAGAATATC TACAAAATTA 540  
ACCTCAGGAA TATTTTTGTT TTTTCAATTG AGGCCATAAT ATACNTNCCG ATNGAAAAAT 600  
TTTNCANCAT ATCNCTAATA TCAAAAAATT ATGATGTTAG TAAACGTAAA AAATTTACAC 660  
40 AAAATAANTT TCACAAACT TANNGGGGAA ATTGGAACAA ANAAAAGACT GGTGAGTGAT 720  
AAGCGATGAT GGCCGGTGAA TCAGGTAGCC GTCCTACAAC GTGGTTGATT TTGAGCAAAC 780  
45 TCCTATCTAC TCTTCACACT ATTGGAAATC CCAAATGTC GTCACACCAT AATAATGTGA 840  
ATTTTGTTAT GGAATTTGAG GGAAACAGTA GATATATGTT TCAACCAGTG AAAGTTACCC 900  
TCCTTTGGAC ATATCTACGA NAGTAGAAAG TAGAAACATT CACTAAACGT GACAACTTTA 960  
50 TAAATTTTCT TTTTGTAAC TTTCTTTAGA TTTATTTACG ANAAGAGAAA TATAAACGTC 1020  
ATGCTAATAA AAAATGCATT ATTTTCTACC ATCTAGCTAG AATATTGATC AAGTCTTCAC 1080  
55 GTTTTTTGTT TATCTCTTCT CTCATAGGCA TGTCCACAAA AGGGTAAGTT TTACTGGTTC 1140  
AAAATATTGC ATGAGTACTA CTAAGCTCGT ATAGTTTGAT CTTACTATCA TTGCGATGAG 1200

	GGTTGTTAGT TTGGAAGAAA TAAGGATTTA TGCAAATGGT AATCATTATG TCTGCTATTT	1260
	AAGAAGTAAA TTATGATGCT TGTTGCGTGA ACATATTAAA TTTGCGAAAA ATAAGCAAGG	1320
5	ATACACGAGA GAAGCTCAGA TATTCACGTA ACGATGTTTC ATCTCTTCTC ATTGAGGAAA	1380
	CATATGGCCA TGATATAGCT AATAAGCCTA CGGGATTGTC NTTTCAACGC CGAATCTACC	1440
10	AAACTGTTCC ATCTCTTATT ATATATAGTT TGGTTATTTA AGTAATTAGA TGCATCATAA	1500
	TCTTTTTTTC TGCCAGTTGT AATGCAGATA AAAATATATT GGTGTTCTA AGGATTGTTC	1560
	AAACGTGCAT GTGTACAAGT TATTATTTAT ATACTTTCAT CTACATGCGA TGC GTTATTT	1620
15	ATAATGATAA AACTAAGATT TTTAGTTAAA TTTAATAAAG AGCTTACGAG CTACAATTAA	1680
	TTAGAAATGG TTGCTCAGAA ATCAGAATAC TATATATGAA AAAAGAAGTT GGTATACTTG	1740
20	AAAAAAGAAA AAAC TACTTG AAAAGATGGT AAAAGATATA GAACGAGTAT ATATCTTACT	1800
	CAAGCACGAT AGAAGTTTGT ATCAAAACAT TGC GTTCCAA ACCAATGTTT GAAGATGGTC	1860
	AAAGGTGCTA CTCATGATGT GGTGCGAAGA AGCTTACGAA AAATCTTGCA ATGAGAGATA	1920
25	ACTTTATGGG CTGCTTGTTT AATATATTGA AAATCATGGT AGACAACACC AAAC TCTCCT	1980
	TTACCAGAAG TCATATTTCC TTAACCTCAG AATAAGTAAA TCTTCTAGTT TATTATTTGA	2040
30	AAGTTGAGCG TATAATTGCA ATGAACTTT TACCAATTCA CCGCCTCCTA ACTGAGTTGT	2100
	TGTATTATCC TATCTCTTTA GCTATCCTTT CTTTGCTCTT GCTCCACCTG CATGTGGCCT	2160
	CTTTATTTAT AATCTCTCTA GATTCTGCTA AAGATGTNTG TTCAAATGG TTTATCTTTA	2220
35	AGGGAAGCAA AGTGAATGGA AACATTTAAA GAAAAAATAA ACTTTTAGCA GAGTTCCATG	2280
	AGATTTTATA CTGATGATAA CTAAAATAAT CTTATATGCG TAAGATTATT TTAGTTCTAA	2340
40	ACTTCATTTT GAAATGAGAG GTCATTGGCC AGGAAAGATT CAATATTGGT TCTTTGTTAA	2400
	TTCTCGTTGG TTTGTTTTTA GTATGGGCTA GATCCAAAAC AGGTCATGGA CTGGGCCGTA	2460
	AACTCTATCC AAAATTCTTC ATGTTTTTCC ATCTTTCAA AATCTTTATC CACCATTCCA	2520
45	TTACTAGGGT GTTGGTTTTA TTTTATTTGT TGATTAATTA TGTATTAGAA AATGTAAAGC	2580
	AATATTCAAT TGTAACATGC ATCATCTAAC ACCAATATCT TGTACTAACC TTTTGTAATT	2640
50	TTCTTATAAA CATTTTAAAA GGCTAATTTA AATAAAAATT ACAATAAACG TGATAACTCA	2700
	CTTTCGTAAC GCATATTTAT TCAAATATAC CAAAATTTAC CATTTTAAGT AAGAGAATCT	2760
	TTTAAAAATT AATTTTCAAT TTCATTAATT AAGAAACAAA GAATTTACTG AAACCTATAT	2820
55	TTTATTAAAT TTTAATAAAA TATATGACTA AAATAACGTC ACGTGAATCT TTCTCAGCCG	2880
	TTCGATAATC GAATACTTTA TTGACTAAGT ATTTATTTAG AAAATTTTAA ACAACACTTA	2940

	ATTTCTAGAA	ACAAAGAGAG	CCTCATATGT	ATAAAAATCT	TCTTCTTATC	TTTCTTTCTT	3000
	TCTTAATAGT	CTTTATTTTT	ACTTAATTAC	TTGGTAATT	TGTGAAAAAC	ACAACCAATG	3060
5	AGAGAAGAGC	AGTTTGACTG	GCCACATAGC	CAATGAGACA	AGCCAATGGG	AAAGAGATAT	3120
	AGAGACCTCG	TAAGAACCGC	TCCTTTGCCA	TTGTATCAT	CTCTCTATAA	AACCACTCAA	3180
10	CCATCAACCT	NTCTTTGCAT	GCAACAAATC	ACTCAAATAA	TTATTTTATA	AAGAACAAAA	3240
	AAAAAAGAC	GGCAGAGAAA	CAATGGAACG	TGGAGCTCCC	TTCTCTCACT	ATCAGCTACC	3300
	CAAATCCATC	TCTGGTAATC	TAAGTGGCTA	TTGTATACA	GTATATACTT	GCCTCCATGT	3360
15	ATATTTATAT	TCTCGTGAAA	AATTGGAGAC	ATGCTTTATG	AATTTTATGA	GACTTTGCAA	3420
	CAACGAACGA	GATGCTTTCT	CTCTAGAAAT	TTAAATTTAG	ATTTGTGAAG	GTTTTGGGAA	3480
20	TGGCCCGGAG	AAGACGATTT	TATATATACA	TGCATGCAAG	AGTTTGATAT	GTATATTGTT	3540
	TCATCATGGC	TGAGTCAAAG	TTTTATCCAA	ATATTTCCAT	GGTGTGGTAT	TAGTTAAACA	3600
	AATCTCTCGT	ATGTGTCATT	GAATATACCC	GTGCATGTAC	CAGGAATGTT	TTTGATTCTA	3660
25	AAAACGTTTT	TTCTTTTGTT	GTAACGGTTG	AGTTTTTTTC	TCGTTTCAA	AACGAGATTC	3720
	TCGTTTGTCT	CTCCCTTGT	CTAAAAACAT	CTACGGTTCA	TGTGATTCAA	AAACACTAAA	3780
30	AAAATATAAA	CTCATTTTTT	TTTAATACTT	AACATTTTAA	CTATATATAT	ATATATATAT	3840
	ATATATATCT	TATACTAGTC	CCAAGTTTTA	GTGTGAGGTT	TTTTTATTCA	AAATCTATCA	3900
	GTACATTTTT	TGGAAAAGAA	CTAAGTGAAA	TTTTCTCCAA	ATTTTCCTTT	TACTATTGAT	3960
35	TTTTTAATTA	CTGGATGTCA	TTAACTTTAA	TCTTTTGATT	CTTTCAACGT	TTACCATTGG	4020
	GAACCTTCAC	ATGAAATAAA	TGTCTACTTT	ATTGAGTCAT	ACCTTCGTCA	ACATAAATTA	4080
40	ATTGATGTTC	TTCTCCAAAT	TTTGAGTTTT	TGGTTTTTCT	AATAATCTTA	ACGAAAGCTT	4140
	TTTGGTATAC	ATGTAAAACG	TAACGGCAAG	AATCTGAACA	GTCTACTCAA	CGGGGTCCAT	4200
	AAGCTAGAA	TGTAGACCCC	ACAAACTTAC	TCTTATCTTA	TTGGTCCGTA	ACTAAGAACG	4260
45	TGTCCCTCTG	ATTCTCTTGT	TTTCTTCTAA	TTAATTCGTA	TCCTACAAAT	TTAATTATCA	4320
	TTTCTACTTC	AACTAATCTT	TTTTTATTTT	CTAAAGATTT	CAATTTCTCT	CTGTATTTTC	4380
50	TATGAACAGA	ATTGAACTTG	GACCAGCACA	GCAACAACCC	AACCCCAATG	ACCAGCTCAG	4440
	TCATAGTAGC	CGGCGCCGGT	GACAAGAACA	ATGGTATCGT	GGTCCAGCAG	CAACCACCAT	4500
	GTGTGGCTCG	TGAGCAAGAC	CAATACATGC	CAATCGCAAA	CGTCATAAGA	ATCATGCGTA	4560
55	AAACCTTACC	GTCTCACGCC	AAAATCTCTG	ACGACGCCAA	AGAAACGATT	CAAGAATGTG	4620
	TCTCCGAGTA	CATCAGCTTC	GTGACCGGTG	AAGCCAACGA	GCGTTGCCAA	CGTGAGCAAC	4680

	GTAAGACCAT	AACTGCTGAA	GATATCCTTT	GGGCTATGAG	CAAGCTTGGG	TTCGATAACT	4740
	ACGTGGACCC	CCTCACCGTG	TTCATTAACC	GGTACCGTGA	GATAGAGACC	GATCGTGGTT	4800
5	CTGCACTTAG	AGGTGAGCCA	CCGTCGTTGA	GACAAACCTA	TGGAGGAAAT	GGTATTGGGT	4860
	TTCACGGCCC	ATCTCATGGC	CTACCTCCTC	CGGGTCCTTA	TGGTTATGGT	ATGTTGGACC	4920
	AATCCATGGT	TATGGGAGGT	GGTCGGTACT	ACCAAAACGG	GTCGTCGGGT	CAAGATGAAT	4980
10	CCAGTGTGG	TGGTGGCTCT	TCGTCTTCCA	TTAACGGAAT	GCCGGCTTTT	GACCATTATG	5040
	GTCAGTATAA	GTGAAGAAGG	AGTTATTCTT	CATTTTATA	TCTATTCAAA	ACATGTGTTT	5100
15	CGATAGATAT	TTTATTTTTA	TGTCTTATCA	ATAACATTTC	TATATAATGT	TGCTTCTTTA	5160
	AGGAAAAGTG	TTGTATGTCA	ATACTTTATG	AGAACTGAT	TTATATATGC	AAATGATTGA	5220
	ATCCAAACTG	TTTTGTGGAT	TAAACTCTAT	GCAACATTAT	ATATTTACAT	GATCTAAAGG	5280
20	TTTTGTAATT	CAAAGCTGT	CATAGTTAGA	AGATAACTAA	ACATTGTAGT	AACCAAGTTT	5340
	AATTTACTTT	TTTGAGTTTA	CATAACTAAC	CAAGCCAAAA	GGTTATAAAA	TCTAAATTCG	5400
25	TTGAGTTGTC	AAACTTCTGA	AGATTGCTAT	CCTCTTTGAG	TTGCTTTCTT	TTGGGTGCTT	5460
	GAGTTTCATT	AGGCTGAGCT	GACTCGTTGC	TCTCTAGTCT	TTCATCTCTG	TCTTTTCCAA	5520
	GGATTCATAA	CGTTGGTCGC	TCTCTGTTTC	TGCCTACACT	TCTTCAAGGG	ATCATTACTG	5580
30	AGGCTAAGAG	TTAAAGACCT	GAACCATGGT	TTTCTGTAAC	TGGTTCAAGT	TCATTCTCCG	5640
	GTTATTGTGT	GGTTATCTTT	CGGTTAGATT	GAAACCCATA	TGTTTGCTCT	GTTTCTTCTA	5700
35	GTTCCAAGTT	TAATTTCCGG	TTATTGTTTG	GCTTTTTTAA	AGTTTTTAAG	GTCTATTCTA	5760
	TGTAAAGACT	ATTCTACGTA	CGTACATTTA	TCGCAAAATT	GAAAGATTAT	AAAAAAAATT	5820
	GAAAGATCCA	AAGGAAACCA	ATAGATTAAA	CTAAAATGTA	GTATCCTTTT	TATCATTTTA	5880
40	GGCTATGTTT	TCTTTTAAGA	AAGCTTTGGT	AGTTAACTCT	GTTTAAAAGA	AAAAAAGAG	5940
	ATGCATAAAT	TAAATTTAAG	TTTCTAGAAC	TTTTGGATAA	ACATATTAAG	CTAAAGAAAT	6000
45	TAAACTAAAG	GGCGTAAATG	CAAGCTTGTT	ATGCGTTATT	GAAAACATTA	CCTCTAAATT	6060
	AAATAGCCCA	ATATTGAAAA	CCTTAAGCTT	CTTTGATCCC	CTTAACCTGT	TTGTCCACCA	6120
	AGTATTAGTT	CATCTCTTAA	CACGGCAACT	CGAAACGGCA	CAATGGACAA	ACATGGTCTT	6180
50	TCAAAAACCA	CTTCCCAATA	CATCCATCGT	CAAACCTCGT	GCCACATGGT	AAGGTCACCA	6240
	CTATTTCTCC	CTTTTCAAAC	TCCTCCAAAC	AAATTGTGCA	CACACTGGCG	TCAGAGTTGG	6300
55	ATTTCTTCTT	ATTATTATAT	ACTTTCCTTG	CCAAACGGTC	AACCACAAAC	TTATTTGCCG	6360
	GTCTAATTAA	CTCGATATTA	TTG3TGGTCT	CATCAAACGA	GTCAATCCGA	GGAGGAGGTG	6420

GAACAATGAC TTTACAGTAC ATGTAACTA ACGTAGCACA AACTGAAGAG TCTACCATAG 6480  
AAATCGACTT ACAGATTCGT TCAGTGAGTT GAGAGTTAGC AATGTCAACA TATTGTTCCG 6540  
5 AGAGCCCTGC TGAGTACAAC CATTCAATCA GTTTTTTCGA GTCATTAGGG TAGGAGGATA 6600  
TGACACCTTC GTAGTCATTG TACGAGAGAA CGAAATTTGG TGGAAGACTA ATTGATGTGT 6660  
10 CCGATCTTCG GGCACCTACG CAGATTTTGA ATGATCCAGC ATCTTGTGAT TTCGGTTTGA 6720  
GGTCTATTTC GCCGCCAAAG GATATTTCCG CTTCCATAGC TATCAAAGAG AAAGAAAAAT 6780  
AGTGAATCCA AGGTTTtaggg TTTCTTTTCT TTGTCTTNCT TATATATAGA GGCGCTAGAT 6840  
15 TGTATTAAGG ATTATACATA TATATAAGTA ATTGCAATTT GTGAGTTTAT CCTTATTCAT 6900  
TTTTAATTTT ATTTACCTTT ATTTAGTTGA TATTGTGTCC TTTTCCTAGG TAGCATTTCC 6960  
TTCCATCTGT GTTAATTATT AGCATTTCCCT TTCCTTTGTC TTATTTGCCT TTATTTTCGTA 7020  
20 GGAAGAAATC CTTTATGNAC CCCATCTTGG CTGAGAACTT GAGATGATTT TAAATCCTCA 7080  
AAAATTATTC AATTTATGAT TTCGAAATG ATATACACTT TATATTTTCT CCTAAAAAAC 7140  
25 CATATTGTAC TAAGAAAAGT AGAAAACCAG ACTTTTAAAT ATGTTAGATT TTAATTGGGT 7200  
TCTTAAAGTG TTTTAGCGTT TNACACCGGT TATTCTCAA AATCCAACT CTATAATTAT 7260  
AGTTTTTAAG TATAAATTAA TCCGGTTGGC CCAATTAGTG GACCGTTTAA AGAGTAGACA 7320  
30 CTTTTTTTTT TATATATCGA CTACCATAAA ACTTTAACGA TTAATATTTT TGGATAATAA 7380  
GCGATCGTTT TGAGGCGTCC CAATTTTTTT TGTTTCTTTT TATATGAGAA ATGGGTTTAA 7440  
35 GAAAACTGC AATTTTGTCC ATAAAGCTAG TCAGAATTCC TGCAGCCCGG GGGATCCACT 7500  
AGTTCTAGAG CGGCCGCCAC CGCGGTGGAG CTCCAATTCG CCCTATAGTG AGTCGTATTA 7560

40 (2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 amino acids

(B) TYPE: amino acid

45 (C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Pro Ile Ala Asn Val Ile

1

5

55

(2) INFORMATION FOR SEQ ID NO:6:



- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 12 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
- |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile | Gln | Glu | Cys | Val | Ser | Glu | Tyr | Ile | Ser | Phe | Val |
| 1   |     |     |     | 5   |     |     |     |     |     | 10  |     |
- 15 (2) INFORMATION FOR SEQ ID NO:7:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- 20 (ii) MOLECULE TYPE: DNA
- 25

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGAATTCAGC AACAAACCCAA CCCCCA

25

5

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

10

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCTCTAGACA TACAACACTT TTCCTTA

27

20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

25

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

35 ATGACCAGCT CAGTCATAGT AGC

23

(2) INFORMATION FOR SEQ ID NO:10:

40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

45

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

50

GCCACACATG GTGGTTGCTG CTG

23

(2) INFORMATION FOR SEQ ID NO:11:

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

41

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAGATAGAGA CCGATCGTGG TTC

23

10

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TCACTTATAC TGACCATAAT GGTC

24

25

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: DNA

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCATAGATGC ACTCGAAATC AGCC

24

40

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

45

(ii) MOLECULE TYPE: DNA

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCTTGGTAAT AATTGTCATT AG

22

55

## (2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 20 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTAAAAACAT CTACGGTTCA

20

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TTTGTGGTTG ACCGTTTGGC

20

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Leu Pro Ile Ala Asn Val Ala  
1 5

## (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 12 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Gln Glu Cys Val Ser Glu Phe Ile Ser Phe Val  
1 5 10

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising a *LEC1* polynucleotide sequence, which polynucleotide sequence specifically hybridizes to SEQ. ID. No. 1 under stringent conditions.  
5
2. The isolated nucleic acid molecule of claim 1, wherein the *LEC1* polynucleotide is between about 100 nucleotides and about 630 nucleotides in length.
- 10 3. The isolated nucleic acid molecule of claim 1, wherein the *LEC1* polynucleotide is SEQ. ID. No. 1.
4. The isolated nucleic acid molecule of claim 1, wherein the *LEC1* polynucleotide encodes a LEC1 polypeptide of between about 50 and about 210 amino acids.  
15
5. The isolated nucleic acid molecule of claim 4, wherein the LEC1 polypeptide has an amino acid sequence as shown in SEQ. ID. No. 2.
6. The isolated nucleic acid molecule of claim 1, further comprising a  
20 plant promoter operably linked to the *LEC1* polynucleotide.
7. The isolated nucleic acid molecule of claim 6, wherein the plant promoter is from a *LEC1* gene.
- 25 8. The isolated nucleic acid of claim 7, wherein the *LEC1* gene is as shown in SEQ. ID. No. 3.
9. The isolated nucleic acid of claim 7, wherein the *LEC1* gene is as shown in SEQ. ID. No. 4.  
30
10. The isolated nucleic acid of claim 7, wherein the *LEC1* polynucleotide is linked to the promoter in an antisense orientation.

11. An isolated nucleic acid molecule comprising a *LEC1* polynucleotide sequence, which polynucleotide sequence encodes LEC1 polypeptide of between about 50 and about 210 amino acids.
- 5 12. The isolated nucleic acid of claim 10, wherein the LEC1 polypeptide has an amino acid sequence as shown in SEQ. ID. No. 2.
13. A transgenic plant comprising an expression cassette containing a plant promoter operably linked to a heterologous *LEC1* polynucleotide that specifically hybridizes to SEQ. ID. No. 1 under stringent conditions.
- 10 14. The transgenic plant of claim 12, wherein the heterologous *LEC1* polynucleotide encodes a LEC1 polypeptide.
- 15 15. The transgenic plant of claim 13, wherein the LEC1 polypeptide is SEQ. ID. No. 2.
16. The transgenic plant of claim 12, wherein the heterologous *LEC1* polynucleotide is linked to the promoter in an antisense orientation.
- 20 17. The transgenic plant of claim 12, wherein the plant promoter is from a *LEC1* gene.
18. The transgenic plant of claim 16, wherein the *LEC1* gene is as shown in SEQ. ID. No. 3.
- 25 19. The transgenic plant of claim 12, which is a member of the genus *Brassica*.
- 30 20. A method of modulating seed development in a plant, the method comprising introducing into the plant an expression cassette containing a plant promoter

operably linked to a heterologous *LEC1* polynucleotide that specifically hybridizes to SEQ. ID. No. 1 under stringent conditions.

21. The method of claim 19, wherein the heterologous *LEC1*  
5 polynucleotide encodes a LEC1 polypeptide.

22. The method of claim 20, wherein the LEC1 polypeptide has an amino acid sequence as shown in SEQ. ID. No. 2.

10 23. The method of claim 19, wherein the heterologous *LEC1* polynucleotide is linked to the promoter in an antisense orientation.

24. The method of claim 19, wherein the heterologous *LEC1* polynucleotide is SEQ. ID. No. 1.

15 25. The method of claim 19, wherein the plant promoter is from a *LEC1* gene.

20 26. The method of claim 19, wherein the *LEC1* gene is as shown in SEQ. ID. No. 3.

27. The method of claim 19, wherein the plant is a member of the genus *Brassica*.

25 28. The method of claim 19, wherein the expression cassette is introduced into the plant through a sexual cross.

29. An isolated nucleic acid molecule comprising a plant promoter that specifically hybridizes to a polynucleotide sequence consisting of nucleotides 1 to 1998 of  
30 SEQ. ID. No. 3.



30. The isolated nucleic acid molecule of claim 28, wherein the plant promoter sequence consists essentially of nucleotides 1 to 1998 of SEQ. ID. No. 3.

5 31. The isolated nucleic acid molecule of claim 28, wherein the plant promoter sequence is a subsequence of SEQ. ID. No. 4.

32. The isolated nucleic acid molecule of claim 28, further comprising a polynucleotide sequence operably linked to the plant promoter sequence.

10 33. The isolated nucleic acid of claim 30, wherein the polynucleotide sequence operably linked to the plant promoter sequence encodes a desired polypeptide.

34. The isolated nucleic acid molecule of claim 28, wherein the polynucleotide sequence is linked to the promoter in an antisense orientation.

15

35. A transgenic plant comprising an expression cassette containing a *LEC1* promoter operably linked to a heterologous polynucleotide sequence, wherein the *LEC1* promoter specifically hybridizes to SEQ. ID. No. 3 under stringent conditions.

20 36. The transgenic plant of claim 33, wherein the polynucleotide sequence encodes a desired polypeptide.

25 37. The transgenic plant of claim 33, wherein the heterologous polynucleotide sequence is linked to the *LEC1* promoter in an antisense orientation.

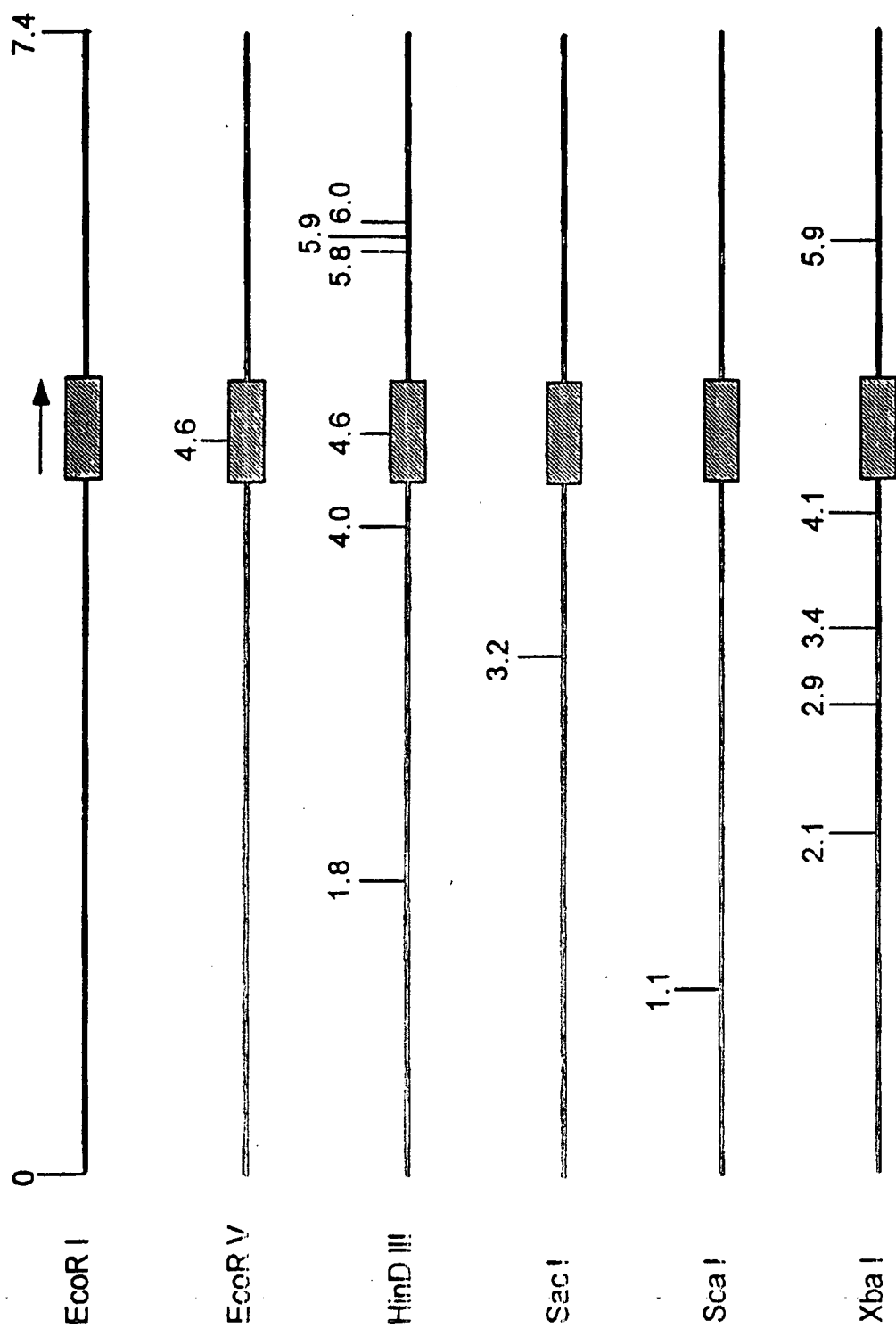
38. The transgenic plant of claim 33, wherein the *LEC1* promoter is as shown in SEQ. ID. No. 3.

30 39. The transgenic plant of claim 33, which is a member of the genus *Brassica*.

40. A method of targeting expression of a polynucleotide to a seed, the method comprising introducing into a plant an expression cassette containing a *LEC1* promoter operably linked to a heterologous polynucleotide sequence, wherein the *LEC1* promoter specifically hybridizes to a polynucleotide sequence consisting of nucleotides 1 to -  
5 1998 of SEQ. ID. No. 3.

41. The method of claim 38, wherein the heterologous polynucleotide sequence encodes a desired polypeptide.

10 42. The method of claim 38, wherein the heterologous polynucleotide sequence is linked to the promoter in an antisense orientation.



Representative restriction sites in a 7.4 kb EcoR I fragment containing the LEC1 gene. Restriction site positions are given in kilobasepairs. The LEC1 protein coding region is between positions 4.4 kb and 5.05 kb and is indicated by the hatched box.

## INTERNATIONAL SEARCH REPORT

 International application No.  
PCT/US98/02998

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC(6) : C12N 15/00, 15/09, 15/29, 15/82; A01H 1/00, 3/00, 5/00 US CL : 536/23.6, 24.1, 24.5; 435/172.3, 252.2, 320.1; 800/205 According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) U.S. : 536/23.6, 24.1, 24.5; 435/172.3, 252.2, 320.1; 800/205 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Agricola, Wpids, STN, Biosis, NCBI, APS		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MEINKE, D.W. A homeotic Mutant of <i>Arabidopsis thaliana</i> with Leafy Cotyledons. Science. 04 December 1992, Vol. 258, pages 1647-1650. See entire reference.	1-42
Y	HECK et al. AGL15, a MADS Domain Protein Expressed in Developing Embryos. The Plant Cell. August 1995, Vol. 7, pages 1271-1282, see entire reference.	1-42
Y	VALVEKENS et al. <i>Agrobacterium tumefaciens</i> -Mediated Transformation of <i>Arabidopsis thaliana</i> Root Explants by Using Kanamycin Selection. Proc. Natl. Acad. Sci., USA, August 1988, Vol. 85, pages 5536-5540, see entire reference	13-28, 35-42
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: *A* document defining the general state of the art which is not considered to be of particular relevance *B* earlier document published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art *A* document member of the same patent family	
Date of the actual completion of the international search 06 APRIL 1998		Date of mailing of the international search report 09 JUN 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230		Authorized officer Ousama M-Faiz Zaghmout Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/02998

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CHAN et al. Novel Gene Expression System for Plant Cells Based on Induction of $\alpha$ -Amylase Promoter by Carbohydrate Starvation. The Journal of Biological Chemistry. 01 July 1994, Vol. 269, No. 26, pages 17635-17641, see entire reference.	7-10, 17, 25, 29-42
Y	SHIMADA et al. Antisense Regulation of the Rice Waxy Gene Expression Using a PCR-Amplified Fragment of the Rice Genome Reduces the Amylose Content in Grain Starch. Theor. Appl. Genet. 1993, Vol. 86, pages 665-672, see entire content.	10, 12, 14, 16-19, 21, 23-28, 34, 37, 42

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/02998

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/02998

### BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not linked as to form a single inventive concept under PCT Rule 13.1.

Group I. Claims 1-6, 11, 13, 15, 20, and 22 are drawn to nucleic acid molecule comprising a LEC1 gene, vectors containing it in sense orientation, methods for their use to transform plants, and the resultant transgenic plants, classified in class 800, subclass 205 for example.

Group II. Claims 7-10, 17, 25, 29- 42 are drawn to LEC1 promoters, methods for their use in constructs to transform plants, and the resultant transgenic plants, classified in class 536, subclass 24.1 for example.

Group III. Claims 10, 12, 14, 16-19, 21, 23-28, 34, 37, and 42 are drawn to vectors containing sequences in antisense orientation, methods for their use to transform plants, and the resultant transgenic plants, classified in class 536, subclass 24.5 for example.

The inventions listed as groups I, II and III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The invention of group I, drawn to a first process of making a first product, involves making the construct and transforming plants not required by the groups II and III.

The invention of group I involves the usage of coding sequences of LEC1 in the sense orientation, not required by groups II-III.

The invention of group II involves the LEC1 promoter sequences and their usage not required by groups I and III, and a non-LEC1 structural gene not required by group I.

The invention of group III involves the usage of coding sequences of LEC1 or other genes in the antisense orientation, and methods for evaluation of gene suppression, not required by groups I-II.